Declaration of Christopher Thompson ISO Plaintiff's Supplemental Sanctions Brief

Redacted Version of Document Sought to be Sealed

BOIES SCHILLER FLEXNER LLP
David Boies (admitted *pro hac vice*)
333 Main Street
Armonk, NY 10504
Tel: (914) 749-8200
dboies@bsfllp.com

Mark C. Mao, CA Bar No. 236165
Beko Reblitz-Richardson, CA Bar No. 238027
Erika Nyborg-Burch, CA Bar No. 342125
44 Montgomery St., 41st Floor
San Francisco, CA 94104
Tel.: (415) 293-6800
mmao@bsfllp.com
brichardson@bsfllp.com
enyborg-burch@bsfllp.com

James Lee (admitted *pro hac vice*)
Rossana Baeza (admitted *pro hac vice*)
100 SE 2nd St., 28th Floor
Miami, FL 33131
Tel.: (305) 539-8400
jlee@bsfllp.com
rbaeza@bsfllp.com

Alison L. Anderson, CA Bar No. 275334
725 S Figueroa St., 31st Floor
Los Angeles, CA 90017
Tel.: (213) 995-5720
alanderson@bsfllp.com

SUSMAN GODFREY L.L.P.
William C. Carmody (admitted *pro hac vice*)
Shawn J. Rabin (admitted *pro hac vice*)
Steven M. Shepard (admitted *pro hac vice*)
Alexander Frawley (admitted *pro hac vice*)
1301 Avenue of the Americas, 32nd Floor
New York, NY 10019
Tel.: (212) 336-8330
bcarmody@susmangodfrey.com
srabin@susmangodfrey.com
sshepard@susmangodfrey.com
afrawley@susmangodfrey.com

Amanda K. Bonn, CA Bar No. 270891
1900 Avenue of the Stars, Suite 1400
Los Angeles, CA 90067
Tel.: (310) 789-3100
abonn@susmangodfrey.com

MORGAN & MORGAN
John A. Yanchunis (admitted *pro hac vice*)
Ryan J. McGee (admitted *pro hac vice*)
201 N. Franklin Street, 7th Floor
Tampa, FL 33602
Tel.: (813) 223-5505
jyanchunis@forthepeople.com
rmcgee@forthepeople.com

Michael F. Ram, CA Bar No. 104805
711 Van Ness Ave, Suite 500
San Francisco, CA 94102
Tel: (415) 358-6913
mram@forthepeople.com

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

CHASOM BROWN, WILLIAM BYATT, JEREMY DAVIS, CHRISTOPHER CASTILLO, and MONIQUE TRUJILLO individually and on behalf of all similarly situated,

Plaintiffs,

vs.

GOOGLE LLC,

Defendant.

Case No.: 4:20-cv-03664-YGR-SVK

DECLARATION OF CHRISTOPHER THOMPSON IN SUPPORT OF PLAINTIFF'S SUPPLEMENTAL SANCTIONS BRIEF

Referral: The Honorable Susan van Keulen

DECLARATION OF CHRISTOPHER THOMPSON

I, Christopher Thompson, declare as follows.

1. I have been retained by Plaintiffs' counsel and asked to provide technical analysis in connection with the Court's Order to Show Cause and Google's Response to that Order to Show Cause.

2. All of the statements in this declaration are true based on my analysis and personal knowledge, and I am available and, if the Court permits it, willing to testify on these matters.

3. A copy of my CV has previously been filed with this Court. Dkt. 536-9. As reflected in my CV, I majored in Computer Engineering and have many years of experience in computing technology. I am being compensated at a rate of $275 per hour for my work in connection with this matter, and none of my compensation is contingent on the outcome of this litigation.

4. In the course of my previous work writing software and building software systems, I designed and used systems that leverage rolling log files. In these systems, logs may be regularly deleted or rotated, depending on the intended use case. For example, log files may be stored in a format more easily accessed for a limited period of time, and then moved to a format that incurs less cost but is harder to access. I also have personal experience working with systems that utilize date information to separate large data sets for efficient query execution.

5. I have also worked as a source code and consulting expert on over 80 matters in which I was tasked with analyzing sometimes voluminous source code productions to identify and describe specific functionality. This source code analysis almost universally occurs without any assistance from the party that developed the code or anyone otherwise knowledgeable about how the system was designed.

6. I have reviewed each and every submission Google and the Special Master made available as part of the Special Master process, including Plaintiffs' data and test data produced by Google, and the transcripts of the hearings before the Special Master. I have also attended several hearings and meet and confers before the Special Master. In addition, all documents Google produced and deposition transcripts for witnesses in this case have been made available to me, pursuant to the Protective Order issued in this case.

7. I was also present at a live test demo with Google engineers and Special Master Douglas Brush on March 4, 2022. At that session, we had tested a small set of Biscotti IDs against three of the "maybe_chrome_incognito" logs.

8. I also testified during the Court's April 21, 2022 sanctions hearing.

Google's Continuing Failure to Identify all Private Browsing Detection Bits and the Logs in which They Have Been Implemented

9. Google states that it has identified [redacted] additional relevant logs via Mr. Sramek's declaration, used by Google to forecast revenue, conduct analysis and testing on users, "join" browsing data with personal information, and track users' interactions with ads related to third-party exchanges.

10. Google's own description of these logs confirms, at a minimum, that Google uses the browsing data it identified as private to monetize on third-party ad exchanges. In addition, Google's description of these logs confirms that it built tools not available to Plaintiffs from the identified private browsing data, including for analysis, testing, and quantification. Mr. Sramek's description of these logs refutes Google's claim that it does not join private browsing data with authenticated data. *See, e.g.*, Srakem Declaration ¶ 9.

11. With the additional maybe_chrome_incongito logs identified by Mr. Sramek, Google has identified it total [redacted] logs with that bit. However, for the is_chrome_incognito and is_chrome_non_incognito_mode bits, Google only identified [redacted] logs, as the "[redacted] logs."

12. As the is_chrome_incognito and is_chrome_non_incognito_mode bits are in the [redacted] proto, a very commonly used proto schema for all Google "upstream logs" (*i.e.*, Google logs that log data at the beginning of the data logging process), it should have been possible for Google to identify all logs containing those bits. Had Google provided access, I could and would have conducted that search.

13. Likewise, the maybe_chrome_incognito bit is contained within the [redacted] proto, which is a different proto commonly used in ads-related "downstream" logs. Notably, Mr. Sramek did not state that Google has identified all logs containing the maybe_chrome_incognito field.

14. The manner in which Google conducted its investigation into identifying incognito detection systems and fields was wholly inadequate. In my experience, analysis of source code resembles a funnel where early, coarse-grained searching gives way to an increasingly refined iterative analysis as the reviewer learns more about the systems at issue.

15. Any source code analysis related to this case necessarily should have begun with the term "incognito" as a search term. Even if that initial search generated a large number of "hits", a conscientious reviewer would review the results, look for patterns, and identify ways to reduce the irrelevant information to improve the quality of the search results.

16. In this case, I would expect that search results would concern two categories: (1) client-side (*e.g.*, Chrome, mobile applications) code, and (2) server-side (*e.g.*, GWS, AdMixer, [redacted]) code. The reviewer could separate the searches and narrow the results relevant to the various heuristics for identifying and/or detecting private browsing activity.

17. For example, Dr. Caitlin Sadowski—Google's Rule 30(b)(6) designated corporate representative for the Incognito-detection bits—testified that she was able to search Google's billions of lines of code for the term "incognito" and used the proposed terms that Plaintiffs identified for the deposition. *See* **Exhibit 8**, Plaintiffs' Notice of Deposition Pursuant to Rule 30(B)(6) (requesting testimony on "Google's development, implementation, and use of any bit or field containing the word 'incognito' or whose name has ever contained the word 'incognito' or whose function was intended to detect Incognito, including the following: is_chrome_incognito, maybe_chrome_incognito, not_chrome_incognito, and chrome_non_incognito. This Topic includes the reasons why Google developed, implemented, and used any such bit or field. This Topic also includes the log or traffic sources as well as the design (including any changes in design) used to determine the bit or field, as well as any logs or data sources where such a bit or field is used and how it is used"). Critically, however, Dr. Sadowski did not conduct her own search, instead only relying on Plaintiffs' proposed search terms.

18. Had Dr. Sadowski performed a more thorough search, she would have readily identified the "[redacted]" field in Google's protos or schema, which could then have identified for Plaintiffs (and the Court) in March 2022, when she was deposed.

19. The first time I learned of the "[redacted]" bit was in December 2022, after Google revealed that field to Plaintiffs' counsel on December 20, 2022, even though Google had identified it no later than October 31, 2022.

20. The shared structure of Google's logs could and should have also informed the analysis and any reasonable investigation by Google. As I stated previously, [redacted] and [redacted] are two known proto structures that contain fields indicative of private browsing activity. While Google represents that it may be challenging to identify every field ever written within a given log, it should be far less challenging to identify the proto used as the base for the log (or any given set of logs).

21. Any reasonably diligent analysis into logs containing these incognito-detection fields should necessarily have also included an identification of *all* logs using those protos as their base. Indeed, Google's response to the pending Order to Show Cause indicated that several of the newly identified logs contained these bits *because they used the* [redacted] *proto*. Dkt. 798 at 8–11.

22. Finally, the supplemental declaration of Eugene Lee suggests that Google has the ability to illustrate the flow of data and fields within Google's logging architecture. I am not aware of Google previously disclosing that ability, which is something I would have found informative as part of this case and with the Special Master process.

The Newly-Disclosed Logs are Accretive

23. I have reviewed Google's representations that the newly-disclosed logs are non-accretive. Google's representations did not include any information about the fields that exist in the logs, and Google's Response did not include any of the schemas for those logs. This is important because without knowing the other fields that exist in these logs, it is impossible to understand and appreciate the full extent of the purposes these logs serve.

24. Google's representations also did not include an exhaustive list of the data sources that the additional logs draw from or later provide data to (which you can think of as upstream and downstream data sources). This is important because these logs contain private browsing data, and any product or service that uses or otherwise relies on the data in these logs would necessarily also

be relying on private browsing data, even if the Incognito-detection bits or pseudonymous identifiers are not communicated to those downstream products or services.

25. Google's representations also did not include an exhaustive list of the uses these logs serve in Google's infrastructure or product development cycles. While Google provided some representations about what these logs are used for, it is unclear whether these logs are available to and used by other products or services in Google.

26. Google's representations about the "[redacted]" logs (Dkt. 798 at 8–11) is particularly concerning, because by their nature (at least from Google's vague description) those logs combine data from different logs to enrich the data with new or different identifiers, properties, and fields that otherwise would not exist in Google's infrastructure.

27. Put simply, Google's "[redacted]" logs make it easier to analyze data because it is in a single place, potentially with additional data sourced from other systems. Moreover, some of these "[redacted]" logs incorporate data from logs that I understand have not been made available to Plaintiffs (*e.g.*, [redacted]) nor have schemas for these logs been provided (cite Hochman Appendix E).

28. I have also reviewed Google's representations concerning preservation. Dkts. 782, 806. I understand that one of Google's primary complaints about its current preservation obligations is that Google has finite, strained storage and resources, and the current preservation obligations are consuming those finite storage and resources. *Id.* The implication of Google's argument is that Google cannot afford to store more than is necessary to support its business functions. Although Plaintiffs were not afforded discovery on this issue, if the newly-disclosed logs at issue in this sanctions briefing were absolutely non-accretive, there would be no purpose for Google to continue storing them with its allegedly finite, strained storage and resources.

29. Since Google's proffered explanations fall short of demonstrating that the newly-disclosed logs are non-accretive, and Google has represented that its finite storage and resources have been strained by the Court's preservation orders, it seems logical that the newly-disclosed logs do serve other purposes for the development, maintenance, and improvement of Google's

products and services. But because Google prevented Plaintiffs from this aspect of discovery during litigation, only Google knows what is going on behind the curtain.

Dr. Psounis's Opinions Relating to the [redacted] Log

30. I have reviewed the declaration filed by Google's expert Dr. Konstantinos Psounis, with Google's Response to the OSC, where Dr. Psounis addresses the [redacted] log. Dkt. 797-21. It is my opinion that Dr. Psounis's opinions about this log are misleading and incomplete.

31. By way of background, this log "contains records from [redacted] logs containing only unauthenticated data" as well as "separate records from [redacted] other logs containing only authenticated data." Google's Response at 10. The [redacted] unauthenticated logs are: [redacted]. The [redacted] authenticated logs are: [redacted]. *See* Panferov Decl. (Dkt. 797-19) ¶ 3.

32. Google did not produce full schema for any of these [redacted] logs in connection with its Response to the OSC, but Google did produce some sample data from [redacted] of these logs during the Special Master process as a part of the Third Iterative Search. Specifically, sample data was produced for [redacted] logs (sample data was not received for the [redacted] log). In addition, Google provided a list of field names from [redacted] of the logs based on a query using [redacted] submitted IDs, and again, data from the [redacted] log was not returned. *See* March 10, 2022 Production, titled "Fields for [redacted] Logs." Google also provided the "top 100 fields" from these [redacted] logs. *See* March 4, 2022 Production, titled "Top 100 fields for [redacted] logs". I have reviewed this test data and list of field names, and based on my review, I can confirm that all [redacted] logs contain the following fields: [redacted].

33. Because all of the logs that comprise the [redacted] log contain the aforementioned fields, it follows that the

[redacted] log, at a minimum, contains these fields. *See* Google's Response to OSC at 9 ("[T]he [redacted] logs contain all of the fields present in their input logs").

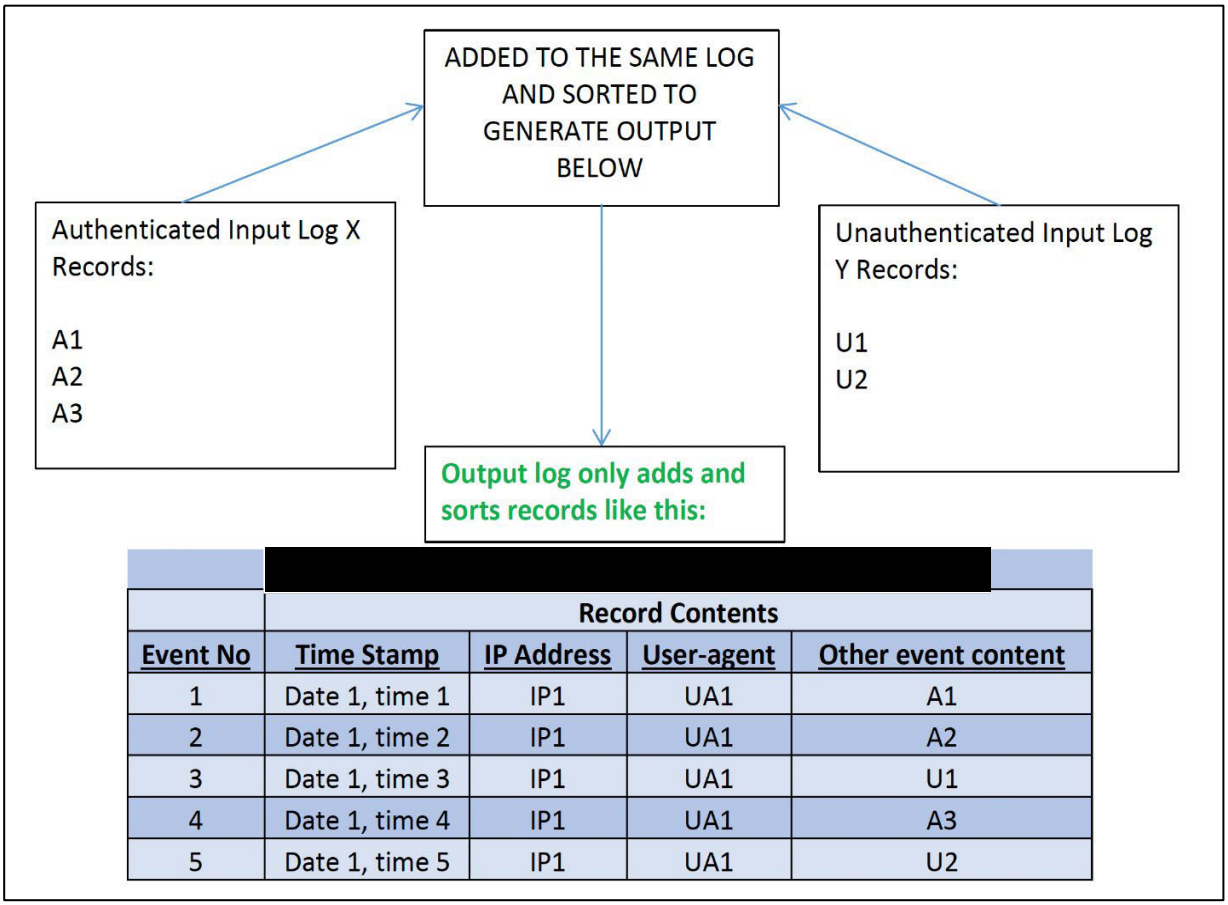
34. Notwithstanding that the log contains both unauthenticated private browsing data as well as authenticated data, Dr. Psounis opines that the [redacted] "does not join authenticated data with unauthenticated data." Psounis Decl. ¶¶ 2, 11. This opinion is misleading and incomplete.

35. Dr. Psounis defines "joining" to mean that "a shared key (or any common data point) was used to associate or combine unauthenticated private browsing data at issue with an individual's Google account." Psounis ¶ 9. Dr. Psounis appears focused on whether the log itself is seeking to associate the data, and he thus glosses over the import of the simple fact that Google is storing authenticated data and unauthenticated data in the same log.

36. Mr. Hochman, Plaintiffs' testifying expert, has proffered opinions in this case about how so-called "unauthenticated" private browsing data can be linked with individuals' Google accounts. For example, in his opening expert report, Mr. Hochman explains how "information tied to a user's Google account could be linked to the same individual's private browsing information stored within Google logs and data sources." Opening Hochman Report (Dkt. 608-12) Opinion 18; *see also id.* Section VIII.F ("Throughout the class period, Google collected and stored private browsing information in ways that can be joined to other Google user information").

37. But it is important to point out that when Mr. Hochman rendered these opinions, Google had not yet revealed that there is a log which contains both "unauthenticated" and "authenticated" data. This new information is significant. In my view, this new information makes it even easier to join private browsing data with users' Google accounts, particularly in light of the fact it is now clear that there is at least [redacted] Google log that contains the following fields for both unauthenticated and authenticated (i.e., GAIA data): [redacted]. Anyone seeking to join the data could easily do so by way of matching data all within the same log.

38. Dr. Psounis prepared a diagram that illustrated the process through which Google combines unauthenticated log entries with authenticated log entries in a single log. *See* Psounis Declaration, Pg. 8. I updated this diagram to illustrate that this Google combination would make identification of class members even easier because of the presence of IP address information, user agent information, and timestamps from both logs. My updated diagram is shown below. Therefore, a user browsing in regular mode who then opened a private browsing window would have data from both contemporaneous sessions in such a way that it would make identifying their traffic through combinations of user agent and IP address trivial.



39. Attached as **Exhibit 9** is a graphic depicting how Google joins unauthenticated and authenticated data.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct. Executed this 20th day of January, 2023, at Nolensville, Tennessee.

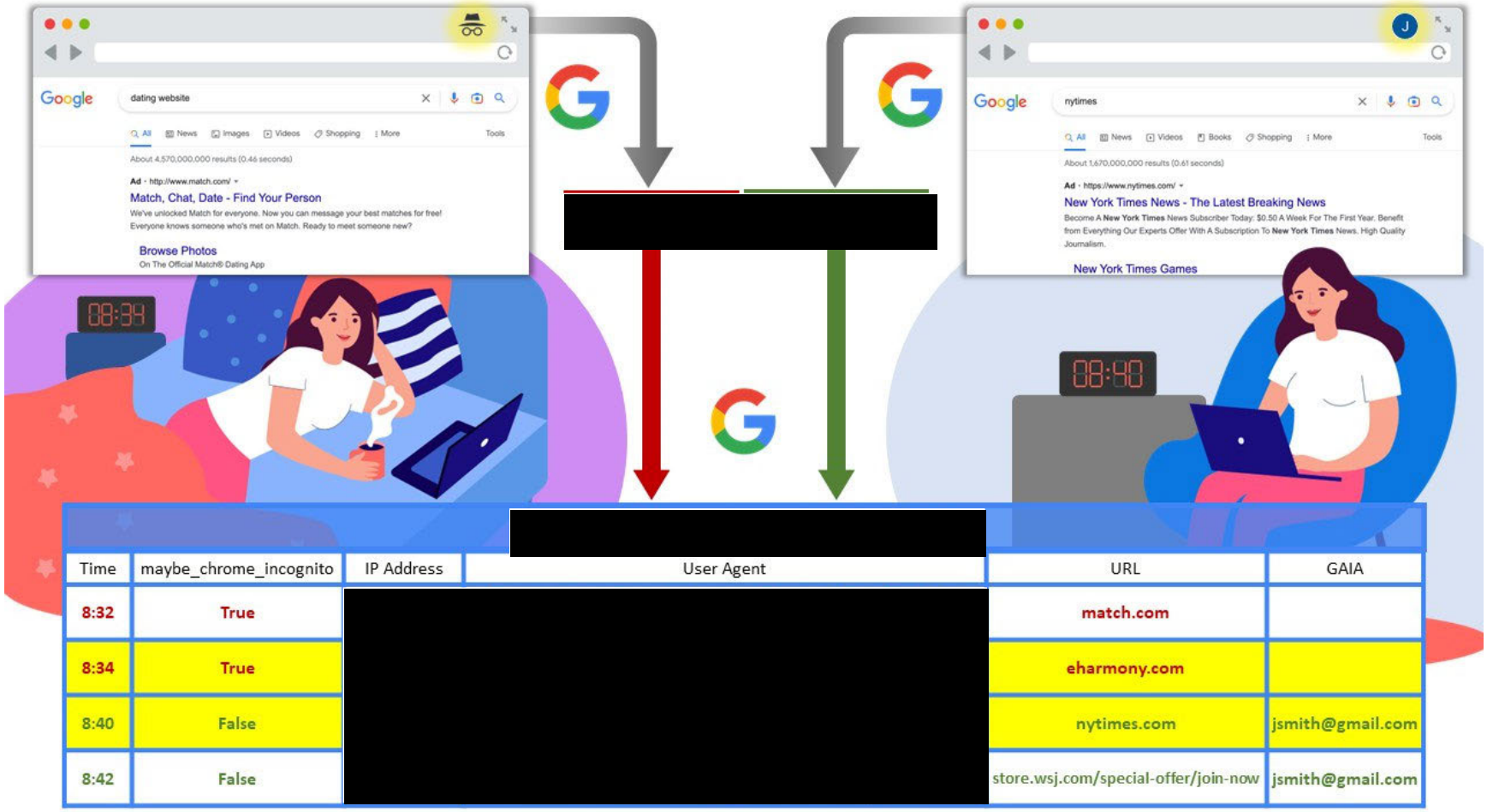
/s/ Christopher Thompson
Christopher Thompson

ATTESTATION

My user ID and password are being used in the electronic filing of this document and, in compliance with N.D. Cal. Civil L.R. 5-1(h)(3), I hereby attest that concurrence in the filing of the document has been obtained from each of the other Signatories.

/s/ Mark C. Mao
Mark C. Mao

EXHIBIT 9



[redacted]

Time	maybe_chrome_incognito	IP Address	User Agent	URL	GAIA
8:32	True	[redacted]		match.com	
8:34	True			eharmony.com	
8:40	False			nytimes.com	jsmith@gmail.com
8:42	False			store.wsj.com/special-offer/join-now	jsmith@gmail.com